Measures of Disease Frequency

Sonya Heltshe, PhD

Assistant Professor, Department of Pediatrics, Seattle Children’s Hospital
How we view the world...

- **Pessimist**: The glass is half empty.

- **Optimist**: The glass is half full.
Epidemiology is...

**Used to**

- identify the cause of disease
- determine the extent of disease
- study the progression of disease
- evaluate preventive and therapeutic measures for a disease or condition
- develop public health policy
• Clinicians are required to know or make estimates of many things:
  – The occurrence of disease in a population
  – The “risk” of developing a disease or an outcome (prognosis)
  – The risks and benefits of a proposed treatment

• This skill requires an understanding of:
  – Measures of disease frequency
    • Case definitions
    • Prevalence and incidence rates
    • Risk (relative and absolute)
Outline

1. Defining a disease ‘case’
2. Measures of disease occurrence
3. Risk estimates and their uses in varying research settings
4. Data sources and Bias
Extent and Progression of Disease

- Who gets the disease?
- In what frequency?
- Is the frequency changing over time?
- How does the frequency in one population compare to the frequency in another population?
- What can disease frequency tell us about associations with exposures or treatment?
1854 London cholera outbreak

Water pump

death
Defining a Disease Case
Case Definition

• Before we can measure disease and disease frequency...
• We must have a **case definition**
  – consistent
  – comparable across populations

<table>
<thead>
<tr>
<th>Age Group</th>
<th>No.</th>
<th>DSM-III</th>
<th>DSM-III-R</th>
<th>DSM-IV</th>
<th>ICD-9</th>
<th>ICD-10</th>
<th>CAMDEX</th>
<th>Clinical Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65–74</td>
<td>391</td>
<td>85 (21.7)</td>
<td>41 (10.5)</td>
<td>43 (11.0)</td>
<td>17 (4.3)</td>
<td>8 (2.0)</td>
<td>7 (1.8)</td>
<td>57 (14.6)</td>
</tr>
<tr>
<td>75–84</td>
<td>931</td>
<td>245 (26.3)</td>
<td>149 (16.0)</td>
<td>114 (12.2)</td>
<td>41 (4.4)</td>
<td>28 (3.0)</td>
<td>49 (5.3)</td>
<td>184 (19.8)</td>
</tr>
<tr>
<td>≥85</td>
<td>557</td>
<td>216 (38.8)</td>
<td>136 (24.4)</td>
<td>100 (18.0)</td>
<td>36 (6.5)</td>
<td>22 (3.9)</td>
<td>36 (6.5)</td>
<td>152 (27.3)</td>
</tr>
<tr>
<td>Total</td>
<td>1879</td>
<td>546 (29.1)</td>
<td>326 (17.3)</td>
<td>257 (13.7)</td>
<td>94 (5.0)</td>
<td>58 (3.1)</td>
<td>92 (4.9)</td>
<td>393 (20.9)</td>
</tr>
</tbody>
</table>

*CSHA denotes the Canadian Study of Health and Aging.

Objectivity and Comparability of Case Definition

Death all causes

Disease specific mortality

Disease - physical signs, lab abnormalities

Disability

Discomfort - symptoms: pain, nausea, cough

Dissatisfaction - emotional reaction: sadness, anger
Lyme Disease Example

- Clinical descriptors
  - ‘bulls eye’ skin lesion > 5cm
  - Joint pain, facial palsy
  - High grade fever

- Lab Criteria
  - *B. burgdorferi* culture
  - Positive IgM antibodies

- Exposure
  - Recall tick bite
  - Presence in county where Lyme disease is endemic

*CDC have very specific criteria for Surveillance of Lyme disease*

*Other studies may define more broadly or specifically*
Specific and (perhaps) well established in your clinical/research area

CDC:
- “Indicators for Chronic Disease Surveillance” *MMWR*, 53(RR-11). Sept 2004
- Infectious disease case definitions
- Chemical poisoning
- Occupational health
- Injury

CDC: Chronic Disease Indicators

The Chronic Disease Indicators (CDI) is a *cross-cutting* set of 97 indicators that were developed by *consensus* and that allows states and territories and large metropolitan areas to *uniformly* define, collect, and report chronic disease data that are important to public health practice and available for states, territories and large metropolitan areas. In addition to providing access to state-specific indicator data, the CDI web site serves as a *gateway* to additional information and data resources. *More...*

Search U.S. Indicator Data

Step 1: Select a state/area:
- United States
- No Comparison
- Selected state's region
- United States
- Alaska
- Arizona
- U.S. Census Regions

Step 2: Compare to:
- (Ctrl + Click to select multiple)
- All Categories

Step 3: Select category of indicators:
- All Categories

Privacy Policy | Accessibility
Home | Contact Us

CDC Home | Search | Health Topics A-Z

This page was last reviewed February, 2012

United States Department of Health and Human Services
Centers for Disease Control and Prevention
National Center for Chronic Disease Prevention and Health Promotion
Metrics of Disease Occurrence
“One’s knowledge of science begins when (s)he can measure what (s)he is speaking about and express it in numbers”

–Lord W.T. Kelvin
Disease Occurrence

• Counts
• Proportion (A/A+B)
  – fraction of population affected
• Rate
  – proportion over a particular time period
    • 14 cases per 1000 per year
• Risk
  – probability of a case in a time period
    • 0.014 cases per person-year
• Rare events
• Large signal

Tetanus – by year, USA, 1955-2000

During 2000, a total of 35 cases of tetanus were reported.

Note: A tetanus vaccine was first available in 1933.
Counts

- Short time periods with little change in time or space

Ebola in West Africa, 2014
Proportion

- **Prevalence** - new and existing cases among population at that time

cases of cold infections in class 4J: Class size = 20

<table>
<thead>
<tr>
<th>January</th>
<th>February</th>
<th>March</th>
</tr>
</thead>
</table>

What was the prevalence February 1st?  
What was the prevalence for all of February?

- Point prevalence = 2/20 = 10%
- Period prevalence = 6/20 = 30%
Proportion and Rate

- **Incidence** -
  - measures *new cases* of the disease or outcome
  - *among those at risk*

- There are two types of incidence measures
  - A proportion
  - A rate
Cumulative Incidence

- proportion of an at-risk group that develops condition over a specific time interval

**Cases of cold infections in class 4J : Class size = 20**

<table>
<thead>
<tr>
<th>January</th>
<th>February</th>
<th>March</th>
</tr>
</thead>
</table>

What is the cumulative incidence in February? 4/18=22.2%
Cumulative Incidence of GI side effects for rofecoxib (VIOXX) vs. naproxen - The VIGOR Trial

Figure 1. Cumulative Incidence of the Primary End Point of a Confirmed Upper Gastrointestinal Event among All Randomized Patients.

Always accompanied by a time period
Incidence as a Rate

**Incidence rate**
- Speed at which an at-risk population develops new disease or condition over time period
- Specifically accounts for time each individual is at risk

**Example**
During a six-month time period, 53 infections were recorded by an infection control nurse at a community hospital. During this time, there were 832 patients with a total of 1,290 patient days.

What is the rate of infections per 100 patient days?

\[
\text{IR} = \frac{53 \times 100}{1,290 \text{ pt. days}} = 4.1 \text{ infections per 100 patient days}
\]
The concept of “person-time”

- the sum of the disease-free time for individuals at risk in the population
  - Remove time WITH disease (not at risk)
  - Loss to follow-up

- 100 people followed for 6 months have same person-time experience as 50 people followed for a year.
  - $100 \times 0.5 = 50$ person-years
  - $50 \times 1.0 = 50$ person-years
The concept of “person-time”

- Calculate by adding up disease-free time
  - 100 subjects followed for 6 months
  - 1 new case develops on day 1 of each successive month
  - Person time = sum of disease free-time for each month (1 thru 6)

  \[
  = 100 + 99 + 98 + 97 + 96 + 95 = 585 \text{ months}
  \]

- Incidence Rate = 5/585 person-months or 8.54 per 1,000 person months
A note on incidence and Recurring events

Number at risk isn’t always exclusively decreasing as patients move in and out of “disease state”
Interpretation of an outcome might differ if a condition can repeatedly occur versus single onset and may depend on your duration of observation.

**Single Event**
- Death
- Rheumatoid arthritis onset
- First pulmonary exacerbation in 2 mos. study window

**Recurring**
- Hospital stay
- Arthritis flare
- Pulmonary exacerbation in 2 year study window

Word of caution:
Determining the end of an event for the purposes of “at risk” population can get messy...
Prevalence and Incidence Rate

- **Prevalence** = **Incidence Rate** X **Disease Duration**

- **High Prevalence** may reflect:
  - High risk
  - Prolonged duration/survival without cure

- **Low Prevalence** may reflect
  - Low risk
  - Short duration
    - Rapid death
    - Rapid cure or recovery
Proportion is not a Rate

Percent of births that were extra-marital in NZ by Year

Gordis, 2000
ProporHon is not a Rate

Births rates by marriage status in NZ by Year

Gordis, 2000
Risk Measures and their uses
Risk and rate are often used interchangeably by epidemiologists but there are differences

- Risk is a probability statement (0-1)
- Risk utilizes a fixed reference period
  - Making it equivalent to cumulative incidence
  - 0.00001 chance of developing cancer in 70 year lifetime
  - 0.02 risk of shunt infection within 30 days of procedure

Rates can be used to estimate risk if time period is short and incidence is relatively constant
Risk and Association

- Risk is a measure of effect used to
  - make comparisons across populations, treatments, or time
  - identify high risk groups/exposures/treatments
  - determine associations and develop hypotheses about disease causality

- Many Risk estimates
  - Relative risk
  - Absolute risk difference
  - Attributable risk

*Presentation and quantification of risk can have a profound impact on clinical decisions*
RotaShield Example

**Intussusception Among Recipients of Rotavirus Vaccine -- United States, 1998-1999**

On August 31, 1998, a tetraivalent live-attenuated rotavirus vaccine (RotaShield™) was licensed in the United States for vaccination of infants. The Advisory Committee on Immunization Practices (ACIP), the American Academy of Pediatrics, and the American Academy of Family Physicians have recommended routine use of RRV-TV for vaccination of healthy infants. During September 1, 1998-July 7, 1999, 15 cases of intussusception (a bowel obstruction in which one segment of bowel becomes

- prospective, observational post marketing surveillance of rotavirus vaccine in US infants

<table>
<thead>
<tr>
<th></th>
<th>No Adverse Events</th>
<th>Intussusception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaccinated</td>
<td>9,789</td>
<td>13</td>
</tr>
<tr>
<td>(N=9802)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No vaccine</td>
<td>48,996</td>
<td>22</td>
</tr>
<tr>
<td>(N=49,000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Risk (V) = 0.00132**

**Risk (NV) = 0.00045**

Baseline risk
Relative Risk (RR)

- The relative probability of the event in the exposed group compared to unexposed group

- $RR = \frac{Risk(V)}{Risk(NV)} = \frac{0.00132}{0.00045} = 2.93$

*Risk of intussusception was ~3 times higher in infants receiving rotavirus vaccine*
Absolute Risk and Attributable Risk (AR)

• The difference in absolute risk (or probability of event) between exposed and un-exposed

• \[ AR = \text{Risk}(V) - \text{Risk}(NV) \]
  
  \[ = 0.00132 - 0.00045 = 0.00087 \]

*Attributable Risk is the absolute risk difference and is the increase in intussusception due to rotavirus vaccine*

Attributable risk depends on baseline risk which can vary markedly across populations
Constant Relative Risk = 3

Adverse Event Rate

Population 1
- Baseline Risk = 20%
- Attributable Risk = 40%
- Treated: 60%
- Untreated: 20%

Population 2
- Baseline Risk = 3%
- Attributable Risk = 6%
- Treated: 9%
- Untreated: 3%

Population 3
- Baseline Risk = 0.16%
- Attributable Risk = 0.36%
- Treated: 0.5%
- Untreated: 0.16%
• Rotavirus causes ~114 million episodes of gastroenteritis
• kills about 1 in 200, or 600,000 children each year
  – Majority in south Asia and sub-Saharan Africa

• 80-100% effective in preventing severe diarrhea in trials in US, Finland, and Venezuela
The RotaShield Story

• Based on intussusception findings, Wyeth withdrew RotaShield from US market 14 months after introduction and halted studies in Bangladesh, India, Ghana and S. Africa

• Never evaluated for risk-benefit in resource poor countries

Attributable Risk for RotaShield is 87 intussusception events per 100,000 vaccinated infants

Vaccinate 100,000 infants at 90% efficacy and 1/200 death rate
Reduction of 450 infant deaths
Framing and Context matter!

- Relative risk is valuable for evaluating strength of association

- Absolute or attributable risk is useful for clinical decision making and public health guidelines

- Effects are generally perceived as more or less favorable when presented in relative terms rather than absolute
Suppose you have a serious disease that needs to be treated with medication. Your risk of dying over the next year is 10% if you don’t receive treatment.

There are only 2 possible medications for the disease: Medication A and Medication B: They cost about the same and have almost no side effects.

Your doctor provides you with the following information about these medications:

Medication A: If you take this medication it will decrease your risk of dying by 80% (four fifths) over the next year.

Medication B: If 100 people with the disease, like you, take this medication 8 deaths can be prevented over the next year.

Question: Which medication do you want? (Circle your answer.)

(1) Medication A
(2) Medication B
(3) Either Medication A or B
(4) Can’t decide

<table>
<thead>
<tr>
<th>Answer</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication A (relative benefit)</td>
<td>56.8</td>
</tr>
<tr>
<td>Medication B (absolute benefit)</td>
<td>14.7</td>
</tr>
<tr>
<td>Medication A or B</td>
<td>15.5</td>
</tr>
<tr>
<td>Can’t decide</td>
<td>13.0</td>
</tr>
</tbody>
</table>

Malenka et al. J Gen Internal Med, 1993
Data Sources and Bias
Bias is a systematic error in the design, conduct or analysis of a study that results in a mistaken estimate of an exposure’s effect on the risk of disease

— (Schlesselman and Stolley, 1982)

**Numerator**

1. Case definition
2. How and where do we find cases?
   - Readily available data
   - Targeted surveys or interviews

**Denominator**

1. Non-representative population
2. Small sample
3. Misclassification
Where do we get our cases?

- Wide reaching, available data registries

Some designed specifically for research:

- Insurance claims data
- Immunization registry (varying quality by location)
- Hospital EMR
Problems with Clinical or Hospital data not intended for research

• Selective with regard to
  – Personal characteristics
  – Disease severity
  – Associated conditions
  – Admission policies

• Error fraught
  – Incomplete
  – Missing
  – Variable diagnostic quality (case definition)

• Undefined risk population (catchment area?)
Where do we get our cases?

• Targeted surveys, studies, interviews
  – Interviewer knows study hypothesis and patient status therefore may probe differently about exposures
  – Subjects may recall exposure in more detail if s/he has the disease (recall bias)
  – Surrogate responders for deceased or children versus source responders
  – Poor response rate
    • e.g. mailed questionnaire
    • Differential response rate = selection bias
Potential Denominator Problems

• Unrepresentative Population /Selection Bias
  – Worker health exposed to occupational hazard compared to general population (healthy worker)
  – Certain populations are undercounted, e.g. refugees, homeless

• Small sample
  – Risk estimates can be very sensitive to small denominators

• Misclassification
  – At risk population may not be easily defined (ethnicity)... similar to problems with case definition
Controlling for data sources and Bias

• **Know** details of any existing data source
  – Intent of data capture
  – Limitations

• **Control** new data collection
  – Sufficient sample size
  – Similar sources and methods across groups
  – Blind interviewers
  – Avoid recall or surrogate responders

• Apply critical evaluation to your own research and that which you review and read
Measures of Disease Frequency

• Defining a disease ‘case’
• Measures of disease occurrence
  – Count, proportion, rate
  – Prevalent versus incident
• Risk estimates
  – explore associations
  – Interpretation and perception
• Data sources and Bias

According to a report released today...

- Smoking
- Exercise
- Fatty Body
- Stress
- Red Wine
- Coffee
- Computer Terminals
- Pancake

Can cause
- Hypothermia
- Heart Disease
- Breast Cancer
- Spontaneous Remission
- Depression
- Glaucoma
- A feeling of well-being
- Wound Healing
- Things

In
- Children
- Two-income families
- Men 25-40
- Overweight
- Smokers
- Rats
- 7 out of 10 women
- Arthritis Sufferers
Questions?